

ANALISIS SENTIMEN APLIKASI PLAYSTORE SIREKAP 2024 PASCA PILPRES DENGAN PERBANDINGAN METODE SUPPORT VECTOR MACHINE (SVM), NAÏVE BAYES CLASSIFIER DAN RANDOM FOREST.

Dede Ardian Tarigan^{*1}, Zakarias Situmorang², Rika Rosnelly³

^{1,3}Universitas Potensi Utama, Medan, ²Universitas Katolik Santo Thomas, Medan
Email: ¹dedeardian2402@yahoo.com, ²zakarias65@yahoo.com, ³rikarosnelly@gmail.com

^{*}Penulis Korespondensi

(Naskah masuk: 18 Desember 2024, diterima untuk diterbitkan: 20 Juli 2025)

Abstrak

Aplikasi Sirekap merupakan sebuah aplikasi berbasis website yang mengandalkan teknologi Optical Character Recognition (OCR) dan Optical Mark Reader (OMR) dalam pengoperasiannya. Perkembangan teknologi ini digunakan untuk mempermudah proses perhitungan suara dengan QuickCount yang sifatnya sementara oleh KPU pada Pemilihan Presiden dan Wakil-Presiden Indonesia Periode 2024. Kedua teknologi tersebut memiliki peran penting dalam mengotomatisasi pola proses baca dan hitung secara real-time. Dengan demikian, analisis sentimen diperlukan untuk mengekstraksi komentar teks dari opini publik tentang aplikasi Sirekap 2024 di Play Store. Penelitian ini berkaitan dengan analisis sentimen terkait hasil perhitungan suara yang menimbulkan ketidaksesuaian di aplikasi Sirekap 2024, apakah bersifat positif atau negatif. Tahapan teknik yang digunakan dalam penelitian ini meliputi scraping data, pre-processing data, pelabelan pola, ekstraksi fitur/pembobotan, pembagian data, dan proses klasifikasi analisis sentimen. Pengumpulan data primer dilakukan menggunakan program Python di aplikasi Google Colab dengan teknik google play scraper di aplikasi playstore android Sirekap 2024. Metode klasifikasi yang digunakan adalah Support Vector Machine, Naïve Bayes, dan Random Forest untuk mengklasifikasikan data. Hasil klasifikasi SVM adalah 82%, Naïve Bayes adalah 71%, dan Random Forest adalah 81%. Dari ketiga metode klasifikasi, kinerja terbaik dalam mengidentifikasi adalah metode klasifikasi SVM dengan akurasi 82%, presisi 82%, recall 82%, dan F1-Score 82%.

Kata kunci: *Sirekap 2024, Support Vector Machine (SVM), Naïve Bayes dan Random Forest.*

SENTIMENT ANALYSIS OF THE SIREKAP 2024 PLAYSTORE APPLICATION POST-PRESIDENTIAL ELECTION WITH COMPARISON OF SUPPORT VECTOR MACHINE (SVM), NAÏVE BAYES CLASSIFIER, AND RANDOM FOREST METHODS.

Abstract

The Sirekap application is a web-based application that relies on Optical Character Recognition (OCR) and Optical Mark Reader (OMR) technology in its operation. The development of this technology is used to facilitate the vote counting process with a temporary QuickCount by the KPU in the 2024 Indonesian Presidential and VicePresidential Elections. Both technologies play an important role in automating the reading and counting process patterns in real-time. Thus, sentiment analysis is necessary to extract text comments from public opinion about the Sirekap 2024 application on the Play Store. This research is related to sentiment analysis concerning the vote count results that cause discrepancies in the Sirekap 2024 application, whether they are positive or negative. The technical stages used in this research include data scraping, data pre-processing, pattern labeling, feature extraction/weighting, data splitting, and the sentiment analysis classification process. Primary data collection was conducted using a Python program in the Google Colab application with the Google Play scraper technique in the Sirekap 2024 Android Play Store application. The classification methods used are Support Vector Machine, Naïve Bayes, and Random Forest to classify the data. The SVM classification result is 82%, Naïve Bayes is 71%, and Random Forest is 81%. Among the three classification methods, the best performance in identification is the SVM classification method with an accuracy of 82%, precision of 82%, recall of 82%, and F1-Score of 82%.

Keywords: *Sirekap 2024, Support Vector Machine (SVM), Naïve Bayes and Random Forest.*

1. PENDAHULUAN

Pemilu merupakan salah satu elemen penting dalam prinsip negara yang berdemokrasi dan berfungsi sebagai instrumen utama dengan dimensi universal guna bertujuan untuk mewujudkan kedaulatan rakyat dalam sistem demokrasi negara. Seiring dengan perkembangan pelaksanaan sistem pemilu, Mahkamah Konstitusi melalui Putusan MK Nomor 14/PUU-XI/2013 memutuskan bahwa Pemilu, baik untuk anggota legislatif maupun presiden dan wakil presiden, harus dilaksanakan secara serentak. Pemilihan Umum (Pemilu) 2024 telah dilaksanakan pada 14 Februari 2024, namun proses demokrasi tersebut tidak lepas dari beberapa permasalahan. Salah satu isu utama yang mendapat perhatian publik adalah adanya ketidaksesuaian data dalam aplikasi Sistem Informasi Rekapitulasi (SIREKAP) yang dikelola oleh Komisi Pemilihan Umum Republik Indonesia (KPU RI). Aplikasi Sistem Informasi Rekapitulasi yang dikenal dengan SIREKAP diluncurkan oleh Komisi Pemilihan Umum Republik Indonesia, merupakan alat pemilu berbasis teknologi yang berkembang untuk menerima dan mengelola penghitungan hasil suara pemilihan umum manual yang diubah menjadi data digital secara real time. Sistem Informasi Rekapitulasi (SIREKAP) menjadi titik awal dalam proses peralihan penyelenggaraan pemilu dari manual ke digital, yang akan menjamin pemilu di Indonesia dapat terselenggara secara efektif dan efisien di masa depan, sehingga pemilu dan demokrasi akan terlihat lebih baik. (Akhsan Firly Saetrian, dkk. 2024).

Pada konteks Pemilu 2024, analisis sentimen publik terhadap hasil Pemilu menjadi kunci untuk memahami pandangan masyarakat, respons terhadap hasil pemilihan, serta dinamika politik yang terjadi. Melalui platform media sosial, masyarakat dapat secara aktif mengekspresikan opini, emosi, dan sentimen terkait hasil Pemilu. (Erik Priansyah dan Tata Sutabri, 2024).

Aplikasi SIREKAP, yang tersedia di Play Store, telah menerima berbagai komentar dan ulasan dari penggunanya. Ulasan tersebut mencerminkan tanggapan pengguna terhadap aplikasi, baik berupa kepuasan maupun kekecewaan atas pengalamannya. (Muhamad Fajar Yudhistira Herjanto dan Carudin, 2024).

Akibat polemik dan kontroversial yang muncul terkait kinerja aplikasi SIREKAP tidak hanya mempengaruhi kepercayaan masyarakat, tetapi juga menjadikan Masyarakat sebagai sasaran bagi para elit politik. Dalam hal ini, suara masyarakat berperan penting sebagai penentu arah kebijakan politik untuk 5 tahun mendatang. Komentar atau opini yang bernuansa positif, netral, maupun negatif, termasuk yang berkaitan dengan isu SARA, Hak Asasi Manusia (HAM), dan ekonomi, berpotensi memicu penyebaran informasi palsu atau keliru. (Gusti Nur Aulia dan Eka Patriya, 2019).

Pada penelitian ini digunakan beberapa penelitian terdahulu sebagai bahan pendukung penelitian. Pada Penelitian yang dilakukan oleh (Gusti Nur, Eka Patriya., 2019). Penelitian ini menerapkan metode Lexicon Based dan Naïve Bayes Classifier untuk menganalisis sentimen pengguna Twitter terkait Pemilihan Presiden (Pilpres) 2019. Proses analisis dimulai dengan mengumpulkan data dari Twitter, diikuti tahap preprocessing data, yang meliputi case_folding atau merubah huruf kecil, normalisasi tweet, filtering, penghapusan stopword atau kata, dan stemming atau merubah menjadi kata baku. Selanjutnya, sentimen diklasifikasikan menggunakan metode Lexicon Based, yang membagi data menjadi 3 kategori yaitu positif, negatif, dan netral. Selain itu, metode Naïve Bayes Classifier juga digunakan untuk klasifikasi sentimen. Penelitian ini bertujuan untuk memberikan wawasan mengenai opini masyarakat di Twitter terkait Pilpres 2019 dalam bentuk sentimen positif, negatif, maupun netral. Hasil klasifikasi menggunakan metode Lexicon Based menunjukkan 484 sentimen positif, 392 negatif, dan 214 netral dari total keseluruhan 1.090 tweet. Sementara itu, metode Naïve Bayes Classifier menghasilkan 206 sentimen positif dan 39 negatif. Pengujian akurasi menggunakan Confusion Matrix menunjukkan bahwa sistem analisis sentimen dengan metode Lexicon Based memiliki tingkat akurasi sebesar 64,49%.

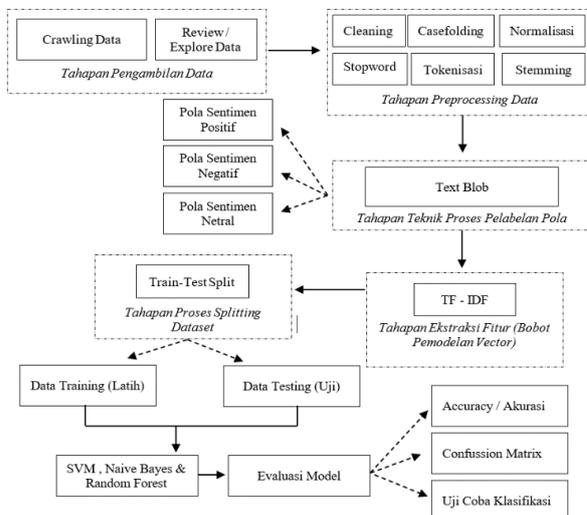
Pada Penelitian yang dilakukan oleh (Nyongki Alexander Radja Bria, Arita Witanti., 2023) Penelitian ini, menganalisis sentimen masyarakat Indonesia terkait Pilpres 2024 menggunakan algoritma Support Vector Machine (SVM). Penelitian ini melibatkan beberapa tahapan, yaitu pengumpulan data (crawling) dari Twitter, pengolahan data, analisis sentimen, preprocessing data, dan evaluasi klasifikasi. Data diambil menggunakan bahasa pemrograman Python melalui aplikasi web Google Colab, dengan memanfaatkan API Key Twitter. Untuk proses klasifikasi, algoritma SVM digunakan untuk mengelompokkan data sentimen. Tahap preprocessing mencakup pembersihan data (cleaning), case_folding, dan tokenizing, yang bertujuan mempersiapkan data sebelum analisis sentimen dilakukan. Selanjutnya, teks data diubah menjadi representasi numerik menggunakan algoritma pembobotan kata seperti Tf-IdfVectorizer. Evaluasi klasifikasi dilakukan dengan membagi data menjadi 80% untuk pelatihan dan 20% untuk pengujian. Hasil evaluasi menunjukkan akurasi sebesar 65%, dengan algoritma SVM lebih unggul dalam mengidentifikasi sentimen positif.

Pada Penelitian yang dilakukan oleh (Gilbert Jeffson Sagala, Yusran Timur Samuel., 2024). Penelitian ini menganalisis sentimen terhadap ulasan aplikasi ChatGPT di Google Play Store dengan menggunakan 3 metode klasifikasi diantaranya metode Random Forest (RF), metode Support Vector Machine (SVM), dan metode Naïve Bayes (NB).

Tujuan penelitian tersebut agar dapat mengevaluasi tingkat akurasi dari masing-masing metode klasifikasi serta melakukan perbandingan antar metode yang digunakan. Tahapan penelitian mencakup pengumpulan dataset, pelabelan data, preprocessing teks, pembobotan term, penerapan metode SMOTE, implementasi algoritma klasifikasi, serta analisis hasil klasifikasi dan evaluasi. Berdasarkan f1-score tertinggi, ketiga metode yang digunakan terbukti efektif. Metode Random Forest menghasilkan akurasi sebesar 91%, presisi positif 92%, recall positif 99%, dan f1-score 90%. Metode SVM juga menunjukkan akurasi sebesar 91%, presisi positif 92%, recall positif 98%, dan f1-score 90%. Sementara itu, metode Naïve Bayes mencatatkan akurasi 90%, presisi positif 90%, recall positif 100%, dan f1-score 87%.

2. METODE PENELITIAN

Penelitian ini akan membahas komentar atau opini pada aplikasi Sirekap KPU di Play Store. Metode perbandingan yang akan digunakan meliputi SVM (Support Vector Machine), Naïve Bayes Classifier, dan Random Forest untuk dilakukan proses ke dalam analisis sentiment. Harapan dari penelitian ini untuk membantu mengklasifikasi komentar atau opini pada aplikasi Sirekap. Hal ini akan memungkinkan untuk memudahkan pengetahuan dan analisis terhadap komentar yang masuk ke dalam kategori klasifikasi pola negatif atau klasifikasi pola positif. Selain itu, penelitian ini juga bertujuan untuk menentukan algoritma mana yang memiliki akurasi terbaik melalui perbandingan ketiga metode yang digunakan. Langkah pertama dalam melakukan penelitian analisis sentimen adalah mengumpulkan dataset. Setelah itu, dataset akan diproses melalui tahap text preprocessing. Selanjutnya, dataset akan dijalani oleh pemodelan klasifikasi yang telah ditetapkan untuk mengetahui hasil uji. (Evita Fitri, dkk. 2020). Berikut Gambar 1 menunjukkan alur kerja dari tahapan penelitian terlampir,



Gambar 1. Tahapan Proses Penelitian.

2.1 Tahapan Pengambilan Data

Pada tahapan ini, penulis mengumpulkan data untuk sebagai dataset yang digunakan melalui scraping web dari ulasan komentar yang ditulis oleh pengguna aplikasi Sirekap2024 di Play Store. Scraping Web adalah proses otomatis dalam mengekstrak data dari halaman website. Proses ini dilakukan dengan menggunakan script atau coding yang membaca konten halaman web dan mengonversikannya menjadi data yang dapat dianalisis, disimpan dan digunakan untuk tujuan penelitian. (Muhamad Fajar Yudhistira Herjanto, dkk. 2024)

2.2 Tahapan Preprocessing Data

Tahapan Preprocessing Data dilakukan untuk membersihkan dan menyiapkan data sebelum dilakukan pembobotan kata dan pelabelan pola sentimen pada tahap selanjutnya. (Dinar Fairus Salsabillah, dkk. 2024). Pada tahapan ini terdapat beberapa langkah,

1. Tahapan Cleaning adalah proses menghapus karakter-karakter yang dianggap tidak penting/perlu.
2. Tahapan Casefolding adalah proses mengubah seluruh teks menjadi huruf kecil.
3. Tahapan Normalisasi adalah pengubahan kata tidak baku menjadi kata baku.
4. Tahapan Stopword Removal adalah proses menghilangkan kata-kata yang dianggap tidak memiliki makna.
5. Tahapan Tokenisasi adalah proses mengubah kalimat menjadi kata per kata.
6. Tahapan Stemming adalah proses menghapus imbuhan awal atau akhiran menjadi kata dasar.

2.3 Tahapan Teknik Proses Pelabelan Pola

Setelah tahap preprocessing dilakukan, langkah berikutnya dengan melabelkan data pada pola. Dalam penelitian ini, teknik TextBlob digunakan untuk melakukan pelabelan pola secara otomatis. TextBlob adalah alat yang menyediakan API untuk pemrosesan Bahasa alami (NLP), yang memungkinkannya melakukan berbagai tugas seperti ekstraksi kata benda, analisis sentiment, klasifikasi, penerjemahan dan lain sebagainya. (Awwaliyah Aliyah, dkk. 2024)

2.4 Tahapan Ekstraksi Fitur

Tahapan Ekstraksi Fitur melibatkan proses identifikasi term atau atribut dalam data, terutama dalam dokumen teks. Proses tahapan ini merupakan bagian penting dalam analisis data, khususnya untuk mengenali istilah atau atribut tertentu dari suatu dokumen. Setiap dokumen teks terdiri atas kata-kata dan proses tokenize yang telah dilakukan tahapan sebelumnya dalam memisahkan dokumen menjadi kata-kata individual atau tunggal. Kemudian, kata-kata atau term tersebut dirubah menjadi nilai statistik yang dapat mudah untuk dihitung. Diantara dari

metode yang paling umum digunakan dalam proses ekstraksi fitur adalah teknik TF-IDF (Term Frequency – Inverse Document Frequency). Teknik metode ini berfungsi untuk mengenali serta mengukur seberapa penting suatu kata dari sebuah dokumen dibandingkan kemunculannya dalam keseluruhan dokumen atau corpus. (Vincentius Riandaru Prasetyo, dkk. 2023)

Berikut rumus dari TF (Term Frequency) adalah

$$TF = \frac{\text{Total Jumlah Dokumen dalam Korpus D}}{\text{Total Jumlah Kata dalam Dokumen D}} \quad (1)$$

Berikut rumus dari IDF (Inverse Document Frequency) adalah,

$$IDF = \log \left(\frac{\text{Total Jumlah Dokumen dalam Corpus}}{\text{Jumlah Dokumen Mengandung Term}} \right) + 1 \quad (2)$$

Maka, dapat dihitung dari perhitungan TF-IDF, berikut rumusnya terlampir,

$$TFIDF(t, d, D) = TF(t, d) \times IDF(t, D) \quad (3)$$

2.5 Tahapan Proses Splitting Dataset

Pada langkah berikutnya, akan dilakukan implementasi pembagian data. Data yang diproses akan dibagi menjadi dua, yaitu data latih (data training) dan data uji (data testing). Kemudian, akan dilanjutkan ke tahap kinerja pembelajaran mesin pada proses data mining. Algoritma Naïve Bayes, SVM, dan Random Forest akan dijalankan untuk mencari nilai akurasi yang akurat dari ketiga metode algoritma tersebut. (Eka ANDrian dan Aulia Rahman Isnain. 2024)

2.6 Tahapan Klasifikasi Metode Algoritma

Dalam tahap pembangunan model klasifikasi, tiga algoritma yang digunakan adalah SVM (Support Vector Machine), Naïve Bayes Classifier, dan Random Forest. Metode ini menggunakan Python sebagai bahasa pemrograman dengan pustaka sklearn dan pipeline. (Putri Elisa dan Aulia Rahman Isnain. 2024). Berikut persamaan rumus dari ketiga metode algoritma yang digunakan,

1 . SVM (Support Vector Machine)

Berikut persamaan rumusnya sebagai berikut,

$$(w \cdot xi + b) \leq 1, yi = -1 \quad (4)$$

$$(w \cdot xi + b) \geq 1, yi = 1 \quad (5)$$

Keterangan:

xi : Representasi data pada posisi ke- i

$w \cdot xi$: Representasi bobot kelas pada data ke- i

b : Acuan nilai bias dalam model

yi : Kelas yang terkait dengan data ke- i

2 . Naïve Bayes Classifier

Berikut persamaan rumusnya sebagai berikut,

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (6)$$

Keterangan:

X : Data pada kelasnya belum diketahui

H : Hipotesis mengenai kelas data

$(H|X)$: Probabilitas hipotesis H terhadap kondisi X

(H) : Probabilitas hipotesis H atau Prior

$(X|H)$: Probabilitas data X terhadap hipotesis H

3 . Random Forest

Berikut persamaan rumusnya sebagai berikut,

$$f(x) = \text{Average}(f_1(x), f_2(x), \dots, f_n(x)) \quad (7)$$

Keterangan:

$f(x)$: Prediksi yang dihasilkan

$f_1 - f_n(x)$: Prediksi setiap pohon keputusan ke- n

(x) : Data yang diberikan sebagai input

2.7 Tahapan Evaluasi Model.

Selama proses penentuan, evaluasi kinerja pada pola sentimen yang digunakan pada penelitian ini dengan visual Confusion Matrix sebagai proses penentuannya. Analisis ini dengan menggunakan Confusion Matrix dapat menghasilkan perhitungan nilai evaluasi model seperti akurasi, presisi, recall dan F1-Score. (M. Irfan Raif, dkk. 2024).

3. HASIL DAN PEMBAHASAN

Berikut adalah hasil dan analisis pembahasan dari setiap tahapan yang dilakukan dalam penelitian ini diantaranya adalah,

3.1 Tahapan Pengambilan Data

Pada proses pengambilan data pada penelitian ini menggunakan Teknik Scraping dengan penerapannya pada Bahasa Pemrograman Python melalui aplikasi website Google Colab dengan teknik google play scraper. Dataset berupa komentar yang akan diambil terdapat pada Aplikasi Sirekap melalui Google Play Store.

Jumlah data komentar yang berhasil dihimpun dari aplikasi sirekap tersebut sebanyak 5000 records, yang dilakukan pengambilannya pada tanggal 05 Juni 2024. Berikut merupakan dataset yang telah berhasil dikumpulkan pada tabel dibawah ini.

A	B	C	D	E	F	G	H	I	J	K	L
reviewId	userName	userImage	content	score	thumbsUp	reviewCreatedAt	replyContent	replyAt	appVersion		
0_35f8099f	Pengguna	https://p/	Aplikasi butut, herannya kpu dika	1	170	2.32	2/26/2024	13:47			2.32
1_eb573c39	Pengguna	https://p/	Segera perbaiki menu detail pemri	1	164	2.41	2/10/2024	11:44			2.41
2_c6caef93	Pengguna	https://p/	APLIKASI GAGAL. Suka nge bug, o	1	514	2.41	2/16/2024	12:29			2.41
3_523ce293	Pengguna	https://p/	Aplikasi belum layak untuk dioper	1	3268	2.3	2/4/2024	17:55			2.3
4_7115cc09	Pengguna	https://p/	Jelek, dari segi tampilan ui nya te	1	0	2.52	9/4/2024	8:19			2.52
5_05417d76	Pengguna	https://p/	Tolong lebih ditingkatkan lagi, ter	2	154	2.41	2/16/2024	16:12			2.41
6_8b563e60	Pengguna	https://p/	Aplikasi sangat tidak layak diguna	1	920	2.41	2/12/2024	1:37			2.41

Gambar 2. Sampel Dataset Awal berupa CSV.

3.2 Tahapan Preprocessing Data

Proses pada pre-processing ini, terdapat beberapa tahapan diantaranya sebagai berikut

1. Proses Cleaning

Pada tahapan ini, semua karakter-karakter seperti tanda baca, emotikon, lambang bendera, dan sebagainya akan dihilangkan atau dihapus serta juga spasi yang berlebihan, URL, hashtag maupun simbol lainnya. Berikut sampel Proses Cleaning yang akan ditampilkan pada gambar 3 dibawah ini,

	content	score	cleaning
0	Aplikasi butut, herannya kpu dikasih waktu 5 t...	1	Aplikasi butut herannya kpu dikasih waktu tah...
1	Segera perbaiki menu detail pemilihan, sebab g...	1	segera perbaiki menu detail pemilihan sebab ga...
2	APLIKASI GAGAL Suka nge bug, over akses bikin L...	1	APLIKASI GAGAL Suka nge bug over akses bikin L...
3	Aplikasi belum layak untuk dioperasikan bisa d...	1	Aplikasi belum layak untuk dioperasikan bisa d...
4	Jelek, dari segi tampilan ui nya terkesan jad...	1	Jelek dari segi tampilan ui nya terkesan jadu...

Gambar 3. Sampel Proses Cleaning.

2. Proses Case Folding

Pada proses tahapan ini, akan mengubah keseluruhan teks menjadi huruf kecil. Berikut sampel Proses Casefolding yang akan ditampilkan pada gambar 4 dibawah ini,

	cleaning	case_folding
	Aplikasi butut herannya kpu dikasih waktu tah...	aplikasi butut herannya kpu dikasih waktu tah...
	Segera perbaiki menu detail pemilihan sebab ga...	segera perbaiki menu detail pemilihan sebab ga...
	APLIKASI GAGAL Suka nge bug over akses bikin L...	aplikasi gagal suka nge bug over akses bikin L...
	Aplikasi belum layak untuk dioperasikan bisa d...	aplikasi belum layak untuk dioperasikan bisa d...
	Jelek dari segi tampilan ui nya terkesan jadu...	jelek dari segi tampilan ui nya terkesan jadu...

Gambar 4. Sampel Proses Case Folding.

3. Proses Normalisasi

Pada proses tahapan ini, perubahan kata tidak baku menjadi kata baku. Berikut sampel Proses Normalisasi yang akan ditampilkan pada gambar 5 dibawah ini,

	case_folding	Normalisasi
	aplikasi butut herannya kpu dikasih waktu tah...	aplikasi butut herannya kpu dikasih waktu tah...
	segera perbaiki menu detail pemilihan sebab ga...	segera perbaiki menu detail pemilihan sebab ga...
	aplikasi gagal suka nge bug over akses bikin L...	aplikasi gagal suka nge bug over akses bikin L...
	aplikasi belum layak untuk dioperasikan bisa d...	aplikasi belum layak untuk dioperasikan bisa d...
	jelek dari segi tampilan ui nya terkesan jadu...	jelek dari segi tampilan ui nya terkesan jadu...

Gambar 5. Sampel Proses Normalisasi.

4. Proses Tokenisasi

Pada proses tahapan ini, proses mengubah kalimat menjadi kata per kata. Berikut sampel Proses Tokenisasi yang akan ditampilkan pada gambar 6 dibawah ini,

	Normalisasi	tokenize
	aplikasi butut herannya kpu dikasih waktu tah...	[aplikasi, butut, herannya, kpu, dikasih, wakt...
	segera perbaiki menu detail pemilihan sebab ga...	[segera, perbaiki, menu, detail, pemilihan, se...
	aplikasi gagal suka nge bug over akses bikin L...	[aplikasi, gagal, suka, nge, bug, over, akses,...
	aplikasi belum layak untuk dioperasikan bisa d...	[aplikasi, belum, layak, untuk, dioperasikan, ...
	jelek dari segi tampilan ui nya terkesan jadu...	[jelek, dari, segi, tampilan, ui, nya, terkesa...

Gambar 6. Sampel Proses Tokenisasi.

5. Proses Stopword Removal

Pada proses tahapan ini, proses menghapus kata-kata yang dianggap tidak ada makna arti. Berikut sampel Proses Stopword Removal yang akan ditampilkan pada gambar 7 dibawah ini,

	tokenize	stopword removal
	[aplikasi, butut, herannya, kpu, dikasih, wakt...	[aplikasi, butut, herannya, kpu, dikasih, ni, ...
	[segera, perbaiki, menu, detail, pemilihan, se...	[perbaiki, menu, detail, pemilihan, gagal, lak...
	[aplikasi, gagal, suka, nge, bug, over, akses,...	[aplikasi, gagal, suka, nge, bug, over, akses,...
	[aplikasi, belum, layak, untuk, dioperasikan, ...	[aplikasi, layak, dioperasikan, komenan, kawan...
	[jelek, dari, segi, tampilan, ui, nya, terkesa...	[jelek, segi, tampilan, ui, terkesan, jadu, b...

Gambar 7. Sampel Proses Stopword Removal.

6. Proses Stemming.

Pada proses tahapan ini, proses menghapus imbuhan awal atau akhiran menjadikannya ke kata dasar. Berikut sampel Proses Stemming yang akan ditampilkan pada gambar 8 dibawah ini

	stopword removal	stemming_data
	[aplikasi, butut, herannya, kpu, dikasih, ni, ...	aplikasi butut heran kpu kasih ni aplikasi has...
	[perbaiki, menu, detail, pemilihan, gagal, lak...	baik menu detail pilih gagal laku koreksi hasi...
	[aplikasi, gagal, suka, nge, bug, over, akses,...	aplikasi gagal suka nge bug over akses lot kam...
	[aplikasi, layak, dioperasikan, komenan, kawan...	aplikasi layak operasi komenan kawanpas unggah...
	[jelek, segi, tampilan, ui, terkesan, jadu, b...	jelek segi tampil ui kes judul banget simulasi...

Gambar 8. Sampel Proses Stemming.

3.3 Tahapan Teknik Proses Pelabelan Pola

Setelah proses tahapan pre-processing telah selesai dilakukan, tahapan berikutnya adalah memberi label pada pola data. Dalam penelitian ini, pelabelan pola dilakukan dengan menganalisa sentimen dari keseluruhan komentar yang terdapat dalam dataset. Sentimen komentar dikategorikan menjadi positif, netral dan negatif berdasarkan pola yang diidentifikasi menggunakan teknik TextBlob. Proses mengenai tahap ini, dengan menerjemahkan komentar Bahasa Indonesia ke Bahasa Inggris terlebih dahulu. Dan kemudian, dilangsungkan pelabelan dengan teknik TextBlob. Proses penerjemah ke Bahasa Inggris ini, penulis menggunakan library googletans versi 3.1. Berikut gambar 9 proses penerjemahannya ke dalam Bahasa Inggris,

```
{'aplikasi butut heran kpu kasih nih aplikasi hasil ada kayak niat hemat budget u
"The application is so outdated, it's surprising that I gave this application, th
```

Gambar 9. Sampel Proses Penerjemah English.

Setelah dilakukan proses terjemahan, kemudian dilanjutkan ke tahapan teknik TextBlob guna proses pembentukan label menjadi 3 pelabelan yang diantaranya positif, negatif dan netral. Berikut gambar 10 hasil proses pelabelan dengan Teknik TextBlob

stemming_data	subjektivitas	polaritas	sentimen
application butut wonder kpu give this applica...	0.866667	-0.910000	negatif
OK, detail menu, select failed, correct the re...	0.266667	0.000000	netral
the application failed like bug over access lo...	0.511111	0.011111	positif
the application is worthy of operation, friend...	0.696667	0.166667	positif
It's ugly in terms of appearance, it's a very ...	0.612500	-0.137500	negatif

Gambar 10. Hasil Pelabelan Teknik TextBlob.

Kemudian pengklasifikasian pola berdasarkan pola sentimen masing-masing serta jumlah kategori diantaranya yaitu sentimen negatif mendapatkan sebanyak 2224, pola sentimen positif mendapat sebanyak 1886 dan sentimen netral sebanyak 872.

sentimen	count
negatif	2224
positif	1886
netral	872

Gambar 11. Hasil Pelabelan Sentimen.

Pencarian pola sentimen pada penelitian ini, hanya berfokus pada pola sentimen negatif dan pola sentiment positif. Sedangkan pola sentimen Netral, tidak diikutsertakan pengujiannya. Dikarenakan pola sentimen netral tidak memberikan kontribusi yang berarti dalam pengambilan Keputusan.

3.4 Tahapan Ekstraksi Fitur

Tahapan ini sering disebut sebagai Count_Vectorizer, yang melibatkan pembobotan-kata atau Term dengan cara menghitung jumlah frekuensi setiap kemunculan kata dalam setiap teks dari keseluruhan dokumen. Hasilnya kemudian direpresentasikan dalam bentuk vector guna mempermudah proses pembelajaran mesin. Berikut gambar 12 merupakan proses pemanggilan dokumen atau teks.,

```
[1448] #memanggil kalimat 1st
stemming_data.tf = stemming_data[3637]
print(stemming_data.tf)

aplikasi mudah ribet loading gabisa login
```

Gambar 12. Proses Ekstraksi Fitur.

Berikut adalah proses perhitungan frekuensi kemunculan kata dalam setiap teks di seluruh dokumen,

DF Scores for Document 3637:

```
login: 1264
loading: 97
mudah: 668
gabisa: 109
ribet: 126
aplikasi: 2386
```

Gambar 13. Proses Perhitungan Frekuensi Kemunculan Term / Kata.

Dan penjelasan dari proses perhitungan TF-IDF terlampir pada tabel 1,

Tabel 1. Proses Perhitungan TF-IDF

TEKS: APLIKASI	TEKS: MUDAH
Word: aplikasi TF: 0.1667 IDF: 1.5434 TF-IDF: 0.2572 DF: 2386 Formula: TF * IDF Calculation: 0.1667 * 1.5434 = 0.2572	Word: mudah TF: 0.1667 IDF: 2.8154 TF-IDF: 0.4692 DF: 668 Formula: TF * IDF Calculation: 0.1667 * 2.8154 = 0.4692
TEKS: RIBET	TEKS: LOADING
Word: ribet TF: 0.1667 IDF: 4.4770 TF-IDF: 0.7462 DF: 126 Formula: TF * IDF Calculation: 0.1667 * 4.4770 = 0.7462	Word: loading TF: 0.1667 IDF: 4.7362 TF-IDF: 0.7894 DF: 97 Formula: TF * IDF Calculation: 0.1667 * 4.7362 = 0.7894
TEKS: GABISA	TEKS: LOGIN
Word: gabisa TF: 0.1667 IDF: 4.6207 TF-IDF: 0.7701 DF: 109 Formula: TF * IDF Calculation: 0.1667 * 4.6207 = 0.7701	Word: login TF: 0.1667 IDF: 2.1784 TF-IDF: 0.3631 DF: 1264 Formula: TF * IDF Calculation: 0.1667 * 2.1784 = 0.3631

3.5 Tahapan Proses Splitting Dataset

Setelah berhasil dilakukan proses ekstraksi fitur atau Count Vectorizer, dilanjutkan proses tahapan splitting dataset. Proses ini dengan membagi Data Training dan Data Testing. Penulis membagi datanya meliputi: 20% Data Testing dan 80% Data Training. Proses pembagian data ini, menggunakan library Sklearn dan Pipeline bertujuan untuk membantu pemrosesan klasifikasi dengan perbandingan 3 metode algoritma yang akan digunakan. Berikut proses Splitting Dataset,

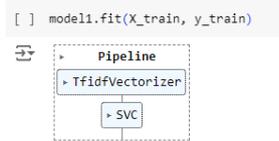
```
[ ] from sklearn.model_selection import train_test_split

# Pembagian data menjadi data pelatihan dan data pengujian
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=120, stratify=y)
```

Gambar 14. Proses Splitting Dataset.

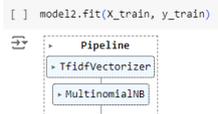
3.6 Tahapan Klasifikasi Metode Algoritma

Pada tahap klasifikasi ini, model dilatih menggunakan metode Support Vector Machine (SVM), Naïve Bayes dan Random Forest dengan data pelatihan yang berasal dari ekstraksi fitur yang telah diproses. Setelah itu, model akan melakukan prediksi akurasi pada data uji. Berikut adalah alur proses dalam klasifikasi menggunakan metode SVM.



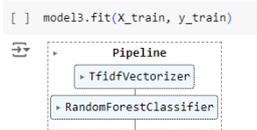
Gambar 15. Proses Pipeline Metode SVM

Berikut proses pipeline pada klasifikasi Metode Naïve Bayes,



Gambar 16. Proses Pipeline Metode Naïve Bayes

Berikut proses pipeline pada klasifikasi Metode Random Forest,

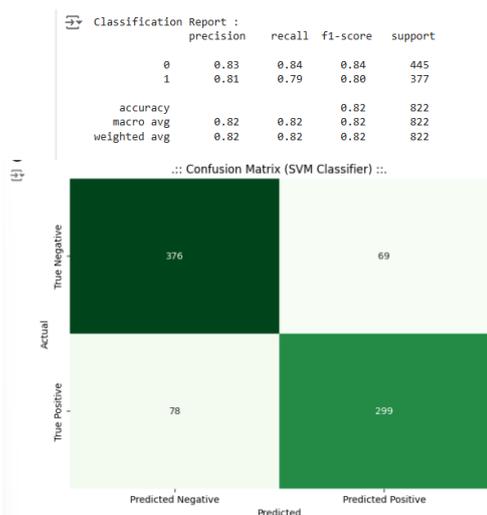


Gambar 17. Proses Pipeline Metode Random Forest

3.7 Tahapan Evaluasi Model.

Proses pada evaluasi model ini merupakan hasil akhir dari tahapan analisis sentimen. Serta model metode algoritma yang digunakan oleh peneliti, diantaranya sebagai berikut,

1. Metode Support Vector Machine (SVM)
Evaluasi model SVM menunjukkan bahwa kelas negatif memiliki persentase precision tertinggi sebesar 83%. Persentase recall tertinggi juga terdapat pada kelas negatif, yaitu 84%. Begitu pula, persentase F1-Score tertinggi tercatat pada kelas negatif, yaitu 84%. Sementara itu, akurasi keseluruhan model SVM mencapai 82%.

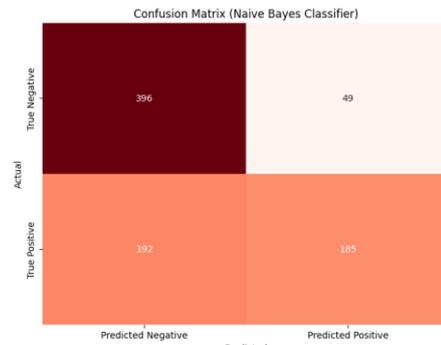


Gambar 19. Evaluasi Akurasi dan Confusion Matrix Metode SVM

2. Metode Naïve Bayes Classifier
Evaluasi model Naïve Bayes menunjukkan bahwa kelas positif memiliki tingkat presisi tertinggi sebesar

79%. Persentase recall tertinggi tercatat pada kelas negatif, yaitu 89%. Adapun, persentase F1-Score tertinggi juga ada pada kelas negatif, yaitu 77%. Sementara itu, akurasi keseluruhan model Naïve Bayes Classifier mencapai 71%.

Classification Report :				
	precision	recall	f1-score	support
0	0.67	0.89	0.77	445
1	0.79	0.49	0.61	377
accuracy			0.71	822
macro avg	0.73	0.69	0.69	822
weighted avg	0.73	0.71	0.69	822

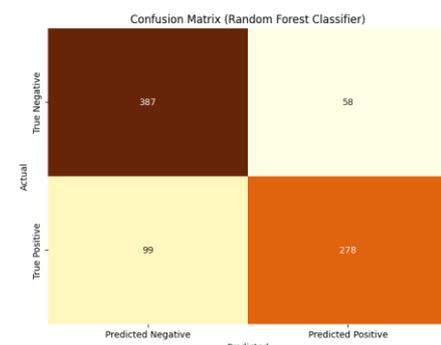


Gambar 18. Evaluasi Akurasi dan Confusion Matrix Metode Naïve Bayes

3. Metode Random Forest

Evaluasi model Random Forest menunjukkan bahwa kelas positif memiliki tingkat akurasi tertinggi sebesar 83%. Persentase recall tertinggi tercatat pada kelas negatif, yaitu 87%. Sedangkan, persentase F1-Score tertinggi juga terdapat pada kelas negatif, yaitu 83%. Sementara itu, akurasi keseluruhan model Random Forest mencapai 81%.

Classification Report :				
	precision	recall	f1-score	support
0	0.80	0.87	0.83	445
1	0.83	0.74	0.78	377
accuracy			0.81	822
macro avg	0.81	0.80	0.81	822
weighted avg	0.81	0.81	0.81	822



Gambar 20. Evaluasi Akurasi dan Confusion Matrix Metode Random Forest

Kata-kata yang banyak muncul dalam bentuk representasi visual dengan Wordcloud pada kelas sentimen negatif terdiri atas “aplikasi”, “udah”, “masuk”, “login” dan “sulit”. Hasil visual dari wordcloud untuk sentimen negatif, dilihat pada gambar berikut 21.



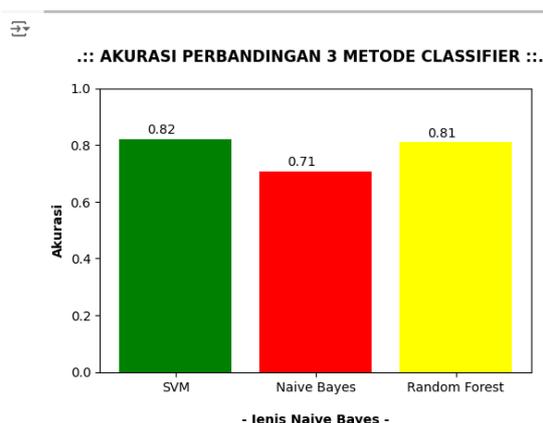
Gambar 21. Visual WordCloud Sentimen Negatif

Sedangkan kata-kata yang banyak muncul dalam bentuk representasi visual dengan Wordcloud pada kelas sentimen positif terdiri atas "aplikasi", "login", "masuk", "baik", dan "bagus". Hasil visual dari wordcloud untuk sentimen positif, dilihat pada gambar 22.



Gambar 22. Visual WordCloud Sentimen Positif

Visual perbandingan model dalam penelitian ini bertujuan untuk menggambarkan tingkat akurasi yang dicapai oleh model SVM, Naïve Bayes Classifier dan Random Forest. Model SVM menunjukkan kinerja terbaik dengan akurasi mencapai 0.82 (82%), diikuti oleh Random Forest yang mencatikan akurasi sebesar 0.81 (81%) Sementara itu, pada model Naïve Bayes Classifier hanya memiliki akurasi sebesar 0.71 (71%), sehingga menjadikan model dengan akurasi terendah diantara ketiganya. Berdasarkan hasil tersebut, model SVM terbukti memiliki performa lebih unggul. Berikut gambar 23 merupakan visualisasi perbandingan akurasi antar model yang digunakan terlampir,



Gambar 23. Visualisasi Perbandingan 3 Metode

4. KESIMPULAN.

Pemilihan umum merupakan salah satu bentuk implementasi prinsip demokrasi yang penting bagi suatu negara, dan dengan berkembangnya sistem pemilu, Mahkamah Konstitusi telah mengatur pelaksanaan Pemilu secara serentak. Penelitian ini

berfokus pada analisis sentimen terkait opini Pasca Pilpres 2024 aplikasi Play Store Sirekap2024, dengan membandingkan tiga metode klasifikasi yaitu Support Vector Machine (SVM), Naïve Bayes Classifier, dan Random Forest. Hasil analisis menunjukkan bahwa metode SVM memiliki akurasi tertinggi di antara ketiga metode tersebut, yaitu sebesar 82%, diikuti oleh Random Forest dengan akurasi 81%, dan Naïve Bayes dengan akurasi 71%. Berdasarkan perbandingan tersebut, SVM merupakan model yang paling efektif dalam mengklasifikasikan sentimen komentar atau opini pada aplikasi Sirekap. Visualisasi akurasi model ini memberikan gambaran yang jelas tentang kinerja masing-masing algoritma dalam analisis sentimen.

DAFTAR PUSTAKA

- ALEXANDER RADJA BRIA, NYONGKI, and ARITA WITANTI. 2024. "Analisis Sentimen Masyarakat Indonesia Menggunakan Algoritma Support Vector Machine Tentang Pilpres 2024." *JATI (Jurnal Mahasiswa Teknik Informatika)* 7(6): 3328–33.
- ALIYAH, AWWALIYAH. 2024. "Analisis Sentimen Twitter Terhadap Tren Penyebaran Informasi Pelaku Kejahatan Menggunakan Algoritma Naives Bayes Mengklasifikasikan Sentimen Dari Data Twitter . Hasil Dari Analisis Ini Diharapkan Dapat." (2): 85–97.
- ANDRIAN, EKA, and AULIA RAHMAN ISNAIN. 2024. "Analisis Sentimen Masyarakat Terhadap Tiktok Shop Di Twitter Menggunakan Metode Naive Bayes Classifier." *Jurnal Media Informatika Budidarma* 8(April): 788–96.
- AULIA, GUSTI NUR, and EKA PATRIYA. 2019. "Implementasi Lexicon Based Dan Naive Bayes Pada Analisis Sentimen Pengguna Twitter Topik Pemilihan Presiden 2019." *Jurnal Ilmiah Informatika Komputer* 24(2): 140–53.
- ELISA, PUTRI, and AULIYA RAHMAN ISNAIN. 2024. "Comparison of Random Forest, Support Vector Machine and Naive Bayes Algorithms to Analyze Sentiment Towards Mental Health Stigma." *Jurnal Teknik Informatika (JUTIF)* 5(1): 321–29.
- FITRI, EVITA. 2020. "Analisis Sentimen Terhadap Aplikasi Ruangguru Menggunakan Algoritma Naive Bayes, Random Forest Dan Support Vector Machine." *Jurnal Transformatika* 18(1): 71.
- HERJANTO, MUHAMAD FAJAR YUDHISTIRA, and CARUDIN CARUDIN. 2024. "Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier." *Jurnal Informatika dan Teknik Elektro Terapan*

- 12(2): 1204–10.
- IRFAN RAIF, M ET AL. 2024. “Otomatisasi Pendeteksi Kata Baku Dan Tidak Baku Pada Data Twitter Berbasis Kbbi.” 11(2): 337–48.
- JEFFSON SAGALA, GILBERT, and YUSRAN TIMUR SAMUEL. 2024. “Sentiment Analysis on ChatGPT App Reviews on Google Play Store Using random Forest Algorithm, Support Vector Machine and Naïve Bayes.” *International Journal of Engineering Business and Social Science* 2(04): 1194–1204.
- MAULIZA, RISHA NUR, and YOANNES ROMANDO SIPAYUNG. 2024. “Penerapan Text Mining Dalam Menganalisis Pendapat Masyarakat Terhadap Pemilu 2024 Pada Media Sosial X Menggunakan Metode Naive Bayes.” *Technomedia Journal* 9(1): 1–16.
- NASIONAL, SEMINAR, TEKNIK ELEKTRO, SISTEM INFORMASI, and TEKNIK INFORMATIKA. 2024. “Analisis Sentimen Terhadap Kecurangan Pemilu Dan SIREKAP Di Twitter Menggunakan Metode Vader Lexicon Dan Naïve Bayes.” : 397–403.
- OKTAVIANA, NATASYA ELDHA, YUITA ARUM SARI, and INDRIATI INDRIATI. 2022. “Analisis Sentimen Terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features Dan Support Vector Machine.” *Jurnal Teknologi Informasi dan Ilmu Komputer* 9(2): 357–62.
- PRAMANA, DEDI, M AFDAL, MUSTAKIM, and INGGIH PERMANA. 2023. “Analisis Sentimen Terhadap Pemindahan Ibu Kota Negara Menggunakan Algoritma Naive Bayes Classifier Dan K-Nearest Neighbors.” *Jurnal Media Informatika Budidarma* 7(3): 1306–14.
- PRASETYO, VINCENTIUS RIANDARU, GATUM ERLANGGA, and DELTA ARDY PRIMA. 2023. “Analisis Sentimen Untuk Identifikasi Bantuan Korban Bencana Alam Berdasarkan Data Di Twitter Menggunakan Metode K-Means Dan Naive Bayes.” *Jurnal Teknologi Informasi dan Ilmu Komputer* 10(5): 1055–62.
- PRIANSYAH, ERIK, and TATA SUTABRI. 2024. “IJM: Indonesian Journal of Multidisciplinary Analisis Sentimen Berbasis Naïve Bayes Pada Media Sosial Twitter Terhadap Hasil Pemilu Indonesia 2024.” *IJM: Indonesian Journal of Multidisciplinary* 2: 128–38.
- SAETRIYAN, AKHSAN FIRLY, INTAN SYAHRANI, MELISA NURDIANA, and MUHAMAD RIZAL FAUZAN. 2024. “Analisis Yuridis Tentang Sengketa Pemilu Terhadap Pelaksanaan Sistem Informasi Rekapitulasi (SIREKAP) Yang Terindikasi ‘ Defect ’ Pada Pemilu Tahun 2024 Yang Berpotensi Merugikan Bakal Calon Presiden.” (3).
- SALSABILLAH, DINAR FAIRUS, DIAN EKA RATNAWATI, and NANANG YUDI SETIAWAN. 2024. “Analisis Sentimen Ulasan Rumah Makan Menggunakan Perbandingan Algoritma *Support Vector Machine* Dengan *Naive Bayes* (Studi Kasus: Ayam Goreng Nelongso Cabang Singosari, Malang).” *Jurnal Teknologi Informasi dan Ilmu Komputer* 11(1): 107–16.

Halaman ini sengaja dikosongkan